Course No. 08 14 S

An introduction to DNA_Stat with applications to input/output manipulation of Next Generation Sequencing data

Next Generation Sequencing (NGS) plays an important role in almost all modern life sciences. Sequencing by NGS technology became so efficient and also so affordable that data analysis has got problems in keeping up with the speed of data generation. Huge files of NGS output have to be stored, parsed and reformatted for subsequent processing by various analysis tools. However, while this activity is easy for computer scientists, it often seriously hinders the work of physicians and biologist who are dependent on software for data reformatting which often either exist nor does work properly. Robert Weissmann's talk, presented at the Berlin Seminar of NGS data analysis early this year, introduced us to the many different formats, read mappers use for output, and the problems we face trying to compare results.

Hence, the offered lectures addresses students and scientists of the
SFB852 who seek independence of computer specialists for data handling.
A very little minimum knowledge of the programming language, C, combined with a powerful special library of functions, called DNA_Stat, allows to analyse, efficiently store and transform data written in any format to any format.

At the end of the course you will agree that DNA_Stat is much easier to learn than EXEL and largely outperforms EXEL in flexibility, speed and data capacity. In addition you will have made your first elementary steps into the large world of C programming that gives you the freedom of using your computer the way you like, no longer being limited to what a given program can make for you. DNA_Stat works under LINUX/UNIX and WINDOWS systems, is freely available and a valuable take home object of the course. The lecturer has used DNA_Stat at many different places of the world, on all kinds of computers including Maqs, and always easily with great success.

Special lecture for the Graduate school

Lecturer: Prof. Paul Wrede, Address: Charité-Universitätsmedizin Berlin, Arnimallee 22, 14195 Berlin Tel 030-8387 1349

Title: Bioinformatic tools for the pattern recognition in amino acid sequences

In the course of many molecular biological investigations a large number of amino acid sequences are generated. Often certain locally encoded signals must be identified in these sequences. For example the identification of the cleavage site for the signal peptidase I in unknown secretory protein precursors is an important task for protein function characterization.

Often sequence alignment of peptides reveals a low identity but in contrast non-homologous sequences can be recognized by the same protein. What is the crucial feature recognized by the protein. Bioinformatic tools as artificial neural networks, support vector machines etc enable the process of feature extraction. Using such a trained feature extraction tool a guided search through the sequence space leads to novel peptide structures.

The students will learn several bioinformatic tools and concepts of rational peptide design and applications in immunology, microbiology as well as molecular biology

PW Seminar: Students give presentations on original publications on the above topics.

Date, time and place for both seminars (Kleffe and Wrede) can be arranged by phone or email:

Kleffe: 838 71310 juergen.kleffe@charite.de

Wrede: 838 71349 paul.wrede@charite.de